TAL1 activation in T-cell acute lymphoblastic leukemia: a novel oncogenic 3' neo-enhancer

Charlotte Smith,[1]* Ashish Goyal,[2]* Dieter Weichenhan,[2] Eric Allemand,[3] Anand Mayakonda,[2] Umut Toprak,[4,5] Anna Riedel,[2] Estelle Balducci,[1] Manisha Manojkumar,[2] Anastasija Pejkovska,[2] Oliver Mücke,[2] Etienne Sollier,[2] Ali Bakr,[2] Kersten Breuer,[2] Pavlo Lutsik,[2] Olivier Hermine,[3,6] Salvatore Spicuglia,[7] Vahid Asnafi,[1] Christoph Plass[2,8] and Aurore Touzart[1,2]

[1]Université de Paris Cité, Institut Necker Enfants-Malades (INEM), Institut National de la Santé et de la Recherche Médicale (Inserm) U1151, and Laboratory of Onco-Hematology, Assistance Publique-Hôpitaux de Paris, Hôpital Necker Enfants-Malades, Paris, France; [2]Division of Cancer Epigenomics, German Cancer Research Center (DKFZ), Heidelberg, Germany; [3]Université de Paris Cité, Institut Imagine, Inserm U1163, Paris, France; [4]Hopp Children's Cancer Center Heidelberg (KiTZ), Heidelberg, Germany; [5]Division of Neuroblastoma Genomics, German Cancer Research Center (DKFZ), Heidelberg, Germany; [6]Department of Hematology, Hôpital Necker Enfants Malades, AP-HP, Faculté de Médecine Paris Descartes, Paris, France; [7]Aix-Marseille University, Inserm, Theories and Approaches of Genomic Complexity (TAGC), Equipe Labellisée Ligue, UMR1090, Marseille, France and [8]German Cancer Research Consortium (DKTK), Heidelberg, Germany

*CS and AG contributed equally as first authors.

Correspondence: A. Touzart
aurore.touzart@aphp.fr

C. Plass
c.plass@dkfz.de

Received: June 16, 2022.
Accepted: December 28, 2022.
Early view: January 12, 2023.

https://doi.org/10.3324/haematol.2022.281583

©2023 Ferrata Storti Foundation
Published under a CC BY-NC license

Supplementary Data

TAL1 activation in T-Cell Acute Lymphoblastic Leukemia: A novel oncogenic 3' Neo-enhancer

Supplementary Methods

Supplementary Table 1-4

Supplementary Figures 1-8

Supplementary References

Supplementary Methods

Patient samples

189 adult and pediatric T-ALL samples were selected based on the availability of DNA/cDNA for 3'NE screening and TAL1 expression analysis. All patient samples had been screened prior for SIL-TAL1 fusion transcripts, TAL1 translocations with TRD and TRB loci and 5'SE mutations. Studies were conducted with informed consent from all patients and in accordance with the Declaration of Helsinki and approved by local and multicenter research ethical committees.

RQ-PCR

Total RNA was harvested using the RNeasy (Qiagen; Hilden, Germany) kit and reverse transcribed with Superscript III (Invitrogen; Waltham, Massachusettes, United States). In cell line models, Quantitative PCR analysis was conducted with the LightCycler® 480-2 system (Roche Professional Diagnostics; Risch-Rokkreuz, Switzerland) using SYBR Green Prima Quant Mix (Steinbrenner Laborsysteme, Wiesenbach, Germany) and specific primers sets for each gene (Supplementary Table 1). Target gene expression was normalized to the housekeeping gene *GAPDH* using the ΔCT method (relative expression is equal to 2-ΔΔCT). In diagnostic and PDX samples, *TAL1* expression was determined using specific primers (Supplementary Table 1), Taqman® probes and the ABI Prism 7500 Sequence Detection System (Applied Biosystems; Waltham, Massachusetts, United States). Gene expression was normalized to the *ABL* and human *GAPDH* (Applied Biosystems) housekeeping genes using the ΔCT method (relative expression is equal to 2-ΔΔCT).

Allelic expression

TAL1 allelic expression was determined by sequencing 9 frequent Single Nucleotide Polymorphisms (SNPs) in the *TAL1* 3´-UTR region in gDNA. Allelic specificity of *TAL1* transcription was then determined by direct sequencing of RT-PCR products from patients with heterozygous SNPs as previously described[1,2].

Blueprint ChIP-seq

Whole total thymus for controls and primary patient cells deriving from blood or bone marrow samples were used for ChIP-seq for H3K4me3, H4K4me1, H3K27ac, H3K27me3, H3K36me3, and H3K9me3 marks as detailed in the Blueprint protocols (https://www.blueprint-epigenome.eu/index.cfm?p=7BF8A4B6-F4FE-861A-2AD57A08D63D0B58).

CRISPR-Cas9 - edited cell lines

To model the patient derived mutation (UPNT-802) in Jurkat and Peer cells, a guide RNA (gRNA) (TAL1 DS crRNA, **Supplementary Table 2**) targeting the 3' region of *TAL1* locus was designed. The gRNA was annealed to equimolar tracrRNA (Alt-R® CRISPR-Cas9 tracrRNA, IDT) by heating to 95°C and slowly cooling to room temperature. 0.5µl of 44µM guide RNA – tracrRNA duplex was complexed with 0.5µl of 36µM Cas9 protein (Alt-R® S.p. HiFi Cas9 Nuclease V3) by mixing and incubating at room temperature for 20 minutes to yield a Cas9 ribonucleoprotein (Cas9 RNP) mix. 2µl of 25µM single stranded oligonucleotide (ssODN) HDR template harboring the mutation and homology to the target site (TAL1 US deletion ssODN) was added to the Cas9 RNP mix along with 25µM electroporation enhancer (Alt-R® Cas9 Electroporation Enhancer). 200 000 Jurkat or Peer cells were washed once with 1X PBS and resuspended in 10µl Neon electroporation Buffer R (Neon transfection system 10µl kit, ThermoFisher Scientific). The Cas9 RNP + ssODN mix was added to the cell suspension and electroporated using Neon transfection system using the following settings: 1625V, 10ms, 3 pulses for Jurkat or 1200V, 40ms, 1 pulse for Peer. 48 hours post-transfection, single cells were seeded onto 96 well plates using serial dilution. Single cell-derived clones were screened for 3'neo-enhancer insertion using 3' Screening F and R primers (**Supplementary Table 1**). Three clones obtained from Peer cells were labelled P-3'NE #1, P-3'NE #2 and P-3'NE #3. One clone obtained from Jurkat cells was labelled J-5'-3'NE. To delete the TAL1 5'super-enhancer in Jurkat and J-5'-3'NE cells, a gRNA (TAL1 US crRNA, **Supplementary Table 2**) targeting the 5'super-enhancer sequence was designed and corresponding Cas9 RNP was prepared as described above. The resulting Cas9 RNPs along with a ssODN HDR template corresponding to the wild type sequence (TAL1 US deletion ssODN) and electroporation

enhancer were electroporated in Jurkat and J-5'-3'NE cells (200,000 cells with following setting: 1625V, 10ms, 3 pulses). Single cell-derived clones were screened for deletion of the TAL1 5'super-enhancer using 5' screening F and R primers (**Supplementary Table 1**). One clone obtained from Jurkat was labelled J-del. Two clones obtained from J-5'-3'NE cells were labelled J-3'NE #1 and J-3'NE #2. All gRNAs were designed using the Custom Alt-R® CRISPR-Cas9 gRNA design tool and purchased from IDT with Alt-R® modification. An ssODN template corresponding to the wild type sequence was designed as per[3] and ordered from IDT (Alt-R™ HDR Donor Oligo).

Cell culture

All commercially available cell lines used within this work were purchased from ATCC collection (ATCC; Manassas, Virginia, United States), mycoplasma free and authenticated. Jurkat, Peer and Molt4 T-ALL cells line were cultured in RPMI medium (Life Technologies; Carlsbad, California, United States) containing 10% (Jurkat and derivative) or 20% (Peer and derivative, Molt4) fetal calf serum, 1% L-glutamine, 100U/ml penicillin-streptomycin (Life Technologies) at 37 °C in the presence of 5% CO_2.

Cell proliferation

Cell number was determined using the automated cell counter Countess (Invitrogen; Waltham, Massachusetts, United States) following Manufacturer instructions. Every 5 days for a total time course of 25 days, culture medium was changed (including doxycycline for J-del + doxy) and total viable cells measured.

Luciferase reporter experiments

A 556 bp genomic fragment around the 3´neo-enhancer mutation, containing either the wild-type sequence (WT) or the mutant allele (MUT), was cloned upstream of luciferase and a minimal promoter in PGL4.23 plasmid. Constructs were electroporated into Jurkat cells. Firefly luciferase activity was measured after 24 hrs, normalized to renilla luciferase in order to control for cell number and transfection efficiency.

H3K27ac and MYB Chromatin Immunoprecipitation

ChIP was performed as per (Bakr et al, NAR 2021, in publication) with minor modifications. 10 million Jurkat, J-3'NE#1, Peer and P-3'NE#1 cells were cross-linked by resuspending cells in 10ml 1% Formaldehyde (Cat.#28906, Thermo Scientific; Waltham, Massachusetts, United States) 1X PBS and incubated for 5 min at 37°C. Glycine was added to a final concentration of 125mM and incubated for 5 min at 37°C. Crosslinked cells were washed with 10ml ice-cold PBS (1000 g, 5 min at 4°C), resuspended in 1ml cell lysis buffer (25mM Tris-HCl, pH=7.4 + 85mM KCl + 0.1% Triton X-100 + 1mM DTT + 1X Halt protease inhibitor), and incubated on a rotator for 30 minutes at 4°C. The nuclei obtained were washed with 500µl of cell lysis buffer (2500g, 5min at 4°C), resuspended in 1ml shearing buffer (10mM Tris-HCl, pH=8 + 0.1% SDS + 1mM EDTA + 1mM DTT + 1X Halt protease inhibitor), and transferred to 1ml AFA milliTUBE with fiber (Cat.#520135, Covaris®). Chromatin was sonicated using Covaris® M220 sonicator (Covaris®; Woburn, Massachusetts, United States) (Peak power=75%, Duty factor=10, Cycles/Burst=200, 15 min) to an average fragment size of 100-500 bp. 5µl of sheared chromatin was de-crosslinked by adding 90µl elution buffer (EB) (10mM Tris-HCl, pH=8 + 5mM EDTA + 300mM NaCl + 0.5% SDS) and 5µl Proteinase K (Qiagen), followed by an incubation for 60 mins at 56°C and then for 30 min at 65°C. DNA was purified using MinElute PCR purification kit (Qiagen) and quantified using Qubit (Thermo Scientific; Waltham, Massachusetts, United States). Sheared chromatin corresponding to 50µg (for H3K27ac ChIP) / 100µg (for MYB ChIP) DNA was diluted to 400µl using ChIP dilution buffer (20mM HEPES + 0.1% SDS + 1% TritonX100 + 150mM NaCl + 1mM EDTA + 0.5mM EGTA). 5µg of H3K27ac antibody (ab4729, Abcam) or 10µg of MYB antibody (Clone 1-1, 05-175 Merck Millipore; Burlington, Massachusetts, United States) was added to the diluted chromatin and incubated on a rotator overnight at 4°C. 100µl Protein A magnetic beads (Cat.# 16-661, Merck Millipore) were washed three times with 1ml 0.5% BSA in 1X PBS and incubated in 250µl 0.5% BSA in 1X PBS on a rotator overnight at 4°C. Beads were resuspended in 100µl ChIP dilution buffer, added to the chromatin-antibody mixture, and incubated on a rotator for 6 hours

at 4°C. Beads were washed twice with 500µl wash buffer 1 (20mM HEPES, 0.1% SDS, 1% TritonX100, 1mM EDTA, 0.5mM EGTA, 150mM NaCl, 0.1% Sodium Deoxycholate), once with 500µl wash buffer 2 (20mM HEPES, 0.1% SDS, 1% TritonX100, 1mM EDTA, 0.5mM EGTA, 500mM NaCl, 0.1% Sodium Deoxycholate), once with 500µl wash buffer 3 (20mM HEPES, 0.1% SDS, 1% TritonX100, 1mM EDTA, 0.5mM EGTA, 250mM LiCl, 0.5% Sodium Deoxycholate, 0.5% NP-40) and then twice with 500µl ice-cold 10mM Tris-HCl, pH=8. Bead bound chromatin was de-crosslinked by resuspending the beads in 200µl EB and 4µl Proteinase K, incubated for 2 hours at 56°C and then for 8 hours at 65°C, followed by addition of 5µl RNAse A (Qiagen) and incubation for 30 min at 37°C. De-crosslinked DNA (in suspension) was eluted using HighPrep™ PCR Clean-up System (Cat.#AC-60050, MAGBIO Genomics; Romanel-sur-Lausanne, Switzerland) as per the manufacturers guidelines. Input and IP genomic DNAs were analyzed by RT–PCR using SybrGreen on a Light Cycler PCR system. IgG control 'cycle over the threshold' Ct values were subtracted from Input or IP Ct values and converted into bound value by $2^{-(\text{IP Ct or input Ct- IgG IP Ct})}$. Allelic-ChIP-qPCR was carried out using allele-specific primers and ChIP-qPCR with primers flanking either the 3'NE or 5'SE. See **Supplementary Table 1** for the primer sequences.

TF binding motifs analysis

Wild-type and mutant 3´neo-enhancer sequences were analyzed using JASPAR.

MYB knockdown using CRISPRi

To knockdown MYB expression in Jurkat and J-3'NE#1 cells, a doxycycline inducible dCas9 KRAB MeCP2 expressing cell line was established. 200 000 cells were washed once with 1X PBS and resuspended in 10µl Neon electroporation Buffer R[4]. 350ng PB-TRE-dCas9-KRAB-MeCP2 and 150ng transposase vector (System Biosciences; Palo Alto; California, United States) were added to the cell suspension and electroporated using the following settings: 1625V, 10ms, 3 pulses. Cells were resuspended in 400µg/ml Hygromycin B (Carl Roth, Germany) containing medium 4 days post-transfection and stably transfected cells were maintained in selection medium. Two independent gRNAs targeting the MYB promoter

(**Supplementary Table 2**), were selected from[5] and cloned into Lenti Guide Puro plasmid as per [4] along with an EGFP targeting gRNA, which served as a non-targeting control. 400 000 HEK293T cells were co-transfected with 1.2µg Lenti Guide Puro – sgRNA-containing plasmid along with 0.9µg psPAX2 and 0.3µg pMD2.G plasmids using 7.5µl TransIT-LT1 (Mirus Bio). 48 hours post-transfection, lentivirus-containing medium was filtered and added to J-3'NE #1 TRE-dCas9-KRAB-MeCP2 cells. 24 hours post-infection cells were resuspended in 1.5µg/ml Puromycin (Sigma-Aldrich, Merck) containing media. Stably transduced cells were maintained in selection media. To induce, dCas9 KRAB MeCP2 expression, cells were resuspended in 2µg/ml Doxycycline (Sigma-Aldrich, Merck; Burlington, Massachusetts, United States) containing medium. 4 days post-doxycycline treatment, cells were harvested for RNA and protein analysis. RNA extraction and qPCR were performed as described in Supplementary methods. Protein was extracted by lysing the cells in 1X RIPA buffer and immunoblotting was performed using antibodies against Cas9 (clone 8C1-F10, Active Motif), MYB (ab45150, Abcam; Cambridge, United Kingdom), TAL1 (clone BTL73, 04-123, Sigma-Aldrich; St. Louis, Missouri, United States) and GAPDH (sc-25778, Santa Cruz Biotechnology Dallas, Texas, United States). PB-TRE-dCas9-KRAB-MeCP2 was a gift from Andrea Califano (Addgene plasmid #122267; http://n2t.net/addgene:122267; RRID: Addgene_122267). lentiGuide-Puro was a gift from Feng Zhang (Addgene plasmid # 52963; http://n2t.net/addgene:52963; RRID: Addgene_52963), (Addgene; Watertown, Massachusetts, United States).

Mi-seq data analysis

Fastq files generated by Mi-seq experiment were aligned to hg19 reference genome with BWA-MEM aligner and processed with SAMtools (http://www.htslib.org) to generate BAM files. In order to phase the reads originating from the same haplotypes, a custom tool was developed which generates the frequency of number of paired-end reads bearing the unique genotypes at the known genetic loci. The tool is written in Rust and available at https://github.com/CompEpigen/hapcounter

ONT sequencing of *TAL1* transcripts

Total RNAs purified from Jurkat cells and PDX cells were reverse transcribed with dT (2/3) and random (1/3) primers and Maxima H Minus Reverse Transcriptase (Thermo Scientific). Prior to ONT sequencing, *TAL1* transcripts were amplified with different primer pairs for 40 amplification cycles using Phusion Plus DNA Polymerase and analyzed by agarose gel (see Figure 5C). For ONT sequencing, each TAL1 amplicon was processed independently in each sample by two PCR steps[6] . First, a pre-amplification of 24-cycles with specific primers corresponding to the first and last exons, and then a second amplification of 20-cycles to add the barcodes. The multiplexed library was constructed using the Oxford Nanopore SQK-LSK109 kit and sequenced on MinION device. Raw data was processed for base calling using "GUPPY [https://community.nanoporetech.com/downloads] and long-read sequences were analyzed with a homemade pipeline. Briefly, using "LAST version 1205 [https://gitlab.com/mcfrith/last]", the sequences were aligned to the Human Genome hg19 and the corresponding BAM files were reviewed manually to annotate all exons. Exons were used to produce a universal reference and all reads were aligned to this particular reference with LAST. Using standard genomic tools, samtools, bedtools, a specific transcript barcode was generated for each sequence depending on its exon architecture, then pooled in clusters defining each alternative *TAL1* transcripts. Given the library preparation, we only retained the sequence containing at least the first and last exons in their barcode. Finally, we only retained transcripts seen 10 times.

Fastq files of ONT sequencing data generated in this study have been deposited in the ENA database (https://www.ebi.ac.uk/ena/browser) under accession code PRJEB56774.

Sanger sequencing: see **Supplementary Table 1** for primer sequences

Supplementary Table 1: Primers sequences

Primer name	Sequence 5'-3'	purpose
TAL1-qPCR-F	CCACCACAATCGAGTGAAGA	RQ-PCR
TAL1-qPCR-R	ACGCCGCACAACTTTGGTG	RQ-PCR
GAPDH-qPCR-F	AGCCACATCGCTCAGACAC	RQ-PCR
GAPDH-qPCR-R	GCCCAATACGACCAAATCC	RQ-PCR
TAL1-Taqman-F	TGGCGGGAGAACAAAGCA	RQ-PCR
TAL1-Taqman-R	CTCTCCATTTTCTCTTCCTACCTCAA	RQ-PCR
Taqman-TAL1-probe-TAMRA	CTATGAGATGGAGATTACTGATG	RQ-PCR
MYB qPCR F	CCTTGCAGCCTTGTAGCAG	RQ-PCR
MYB qPCR R	CTCACATGACCAGCGTCC	RQ-PCR
LMO1 qCPR F	CAGCCACTCTACCAGTGCAG	RQ-PCR
LMO1 qCPR R	AGGAGGGGTCACACACAGA	RQ-PCR
3'NE-F	CACCACACCCGCTAGCCCATGGATGCTCATCCTCC	sanger sequencing
3'NE-R	CACACCCACCCAAGCTTGCCTAGAGGTGAGGGACC	sanger sequencing
5'SE-F	TACCTGTGCACAGCTGGAG	sanger sequencing
5'SE-R	GCTCTCCTGATTAGCATACC	sanger sequencing
3'NE-spe-F	GGATGTGATCCGTTCGCGC	sanger sequencing
3'NE-spe-R	CCCCACTTCCTCTCTTGGG	sanger sequencing
ACTB-CHIP-F	CGCACAGTGCAGCATTTTT	ChIP-Q-PCR
ACTB-CHIP-R	GCCAACGCCAAAACTCTC	ChIP-Q-PCR
3'NE-CHIP-F	AGAGCCTAGGGGAAGAGGC	allele-specific ChIP-Q-PCR
3'NE-wt-CHIP-R	GATGTGATCCGTTCGAAGAG	allele-specific ChIP-Q-PCR
3'NE-mut-CHIP-R	CGCGCTGAACAGACGGTTTAC	allele-specific ChIP-Q-PCR
5'SE-ChIP-qPCR F	TGCCCCACCCCATTCCTATT	ChIP-Q-PCR
5'SE- ChIP-qPCR R	GCACACAGGGCACAAAAAAGG	ChIP-Q-PCR
3'NE- ChIP-qPCR F	CTCTTCCTACCTCTCCCCAC	ChIP-Q-PCR
3'NE- ChIP-qPCR R	GGGAGGAGTGGAGAGTCAGA	ChIP-Q-PCR
3'NE-a-cas-F	CACCACACCCAAGCTTCTAAGACCCAAACATATGCACAT	alpha-cas- PCR amplification
3'NE-a-cas-R	CACACCCACCCACTAGTGCCCGTCCTCCCCACTTCC	alpha-cas- PCR amplification
3'NE-a-cas-mi-seq-i5-F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTAAGACCCAAACATATGCACAT	alpha-cas- PCR amplification
3'NE-a-cas-mi-seq-i7-R	AGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCCCGTCCTCCCCACTTCC	alpha-cas- PCR amplification
5'SE-a-cas-F	CACCACACCCAAGCTTCAGATAAACTGAGGGTCACAG	alpha-cas- PCR amplification
5'SE-a-cas-R	CACACCCACCCACTAGTTGCACAACTGAGGGGTGTTTAAG	alpha-cas- PCR amplification

5’SE-a-cas-mi-seq-F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGATAAACTGAGGGTCACAG	alpha-cas- PCR amplification
5’SE-a-cas-mi-seq-R	AGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGCACAACTGAGGGGTGTTTAAG	alpha-cas- PCR amplification
LMO1-a-cas-F	GGCAGCTAGCGGGCTCTAAT	alpha-cas- PCR amplification
LMO1-a-cas-R	CACCCTACTGCGGAACGGAT	alpha-cas- PCR amplification
LMO1-a-cas-mi-seq-i5-F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCAGTCTCACCTACTTTTGTGCC	alpha-cas- PCR amplification
LMO1-a-cas-mi-seq-i7-R	AGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACAGATCCTGCAGGTGCAGG	alpha-cas- PCR amplification
LMO2-a-cas-F	GAAGATGCCATGGAGACGGCG	alpha-cas- PCR amplification
LMO2-a-cas-R	TCCGAGAACACCTAGTGGTTGCCT	alpha-cas- PCR amplification
LMO2-a-cas-mi-seq-i5-F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTTCTGACAGGCGGCGC	alpha-cas- PCR amplification
LMO2-a-cas-mi-seq-i7-R	AGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAAAGGGGTCTTTGAGATGTCACAC	alpha-cas- PCR amplification
LMO1-FL-cloning-F	GGCAGCTAGCGGGCTCTAAT	cloning
LMO1-FL-cloning-R	CACCCTACTGCGGAACGGAT	cloning
M13 reverse primer	CAGGAAACAGCTATGAC	sequencing
TAL1 - S1	GCAGTCGAACATGTAGCTGACTCAGGTCACTCGTTGCGTACGATTGTGCTCC	ONT sequencing
TAL1 - S3	GCAGTCGAACATGTAGCTGACTCAGGTCACTCGCAGTGACCCCCAGC	ONT sequencing
TAL1 - R2	TGGATCACTTGTGCAAGCATCACATCGTAGGACCAGACCATCAGCAAACAGACATC	ONT sequencing
TAL1 - S9b	GCAGTCGAACATGTAGCTGACTCAGGTCACGAGTTTGGGATCATCAGGCCCAC	ONT sequencing
TAL1 - R5	TGGATCACTTGTGCAAGCATCACATCGTAGGAGACTCAGCTCTGCTCTTCTCTGG	ONT sequencing

Supplementary Table 2: CRISPR-Cas9 cell lines edition. sgRNA and repair templates (ssODN) sequences.

Name	Sequence	purpose
MYB sgRNA#1	GCGCAGCCGGGGAGGGACGC	CRISPRi
MYB sgRNA#2	GAAACTTCGCCCCAGCGGTG	CRISPRi
TAL1-5'-crRNA	AGGGTCACAGAAAGACGGTT**AGG**	5'SE deletion
TAL1-5'-deletion ssODN	CAGGGCACAAAAAAGGATCTGTAGACAAGGGAGGAACTGAATTAATGGTATTTGAAAAAGTGACAGAGACATCTGCCAGGAAGTAGGGTTACGTCTTTCTGTGACCCTCAGTTTATCTGTAATAGGA	
TAL1 3'-crRNA	GATGTGATCCGTTCGAAGAG**AGG**	3'NE introduction
TAL1 3'-insertion ssODN	GGCAGCTGCAGGGGCTGGTCCAAAGGGGAGGGGAGGAGTGGAGAGTCAGAAACGCGCTATCAGCACTGCCCGTCCTCCCCACTTCCTCTCTTGGGTAAACCGTCTGTTCAGCGCGAACGGATCACATCCTGTTATTGTGTGGAAAGAAA	
LMO1 SNP DS crRNA	GGCAAATTGAGCCATTTAGA**AGG**	crRNAs for phasing LMO1 activating mutation in Jurkat cells
LMO1 MUT US crRNA	GGCACTCCGTAGCTGCGGGA**TGG**	
LMO2 SNP DS crRNA	GGCACGAATCCGCTTGTCAC**AGG**	crRNAs for phasing LMO2 activating mutation in MOLT4 cells
LMO2 MUT US crRNA	AGTGGGTCTAGCCTGCAAGC**TGG**	
TAL1 SNP DS crRNA	TGGAATCTTTAGCGCTCAAT**AGG**	crRNAs for phasing TAL1 activating 5'SE in Jurkat cells
TAL1 5'SE US crRNA	ATGAGTTAGACTGTAACGGA**GGG**	
TAL1 3'SE DS crRNA	GAGCTCGGACCAAGCAAGTG**GGG**	crRNAs for phasing TAL1 activating 3'SE in J-3'NE#1 cells
TAL1 SNP US crRNA	TTGGTCATCAAGGGTGCGTC**TGG**	

Supplementary Table 3: ACT-seq oligonucleotides sequences

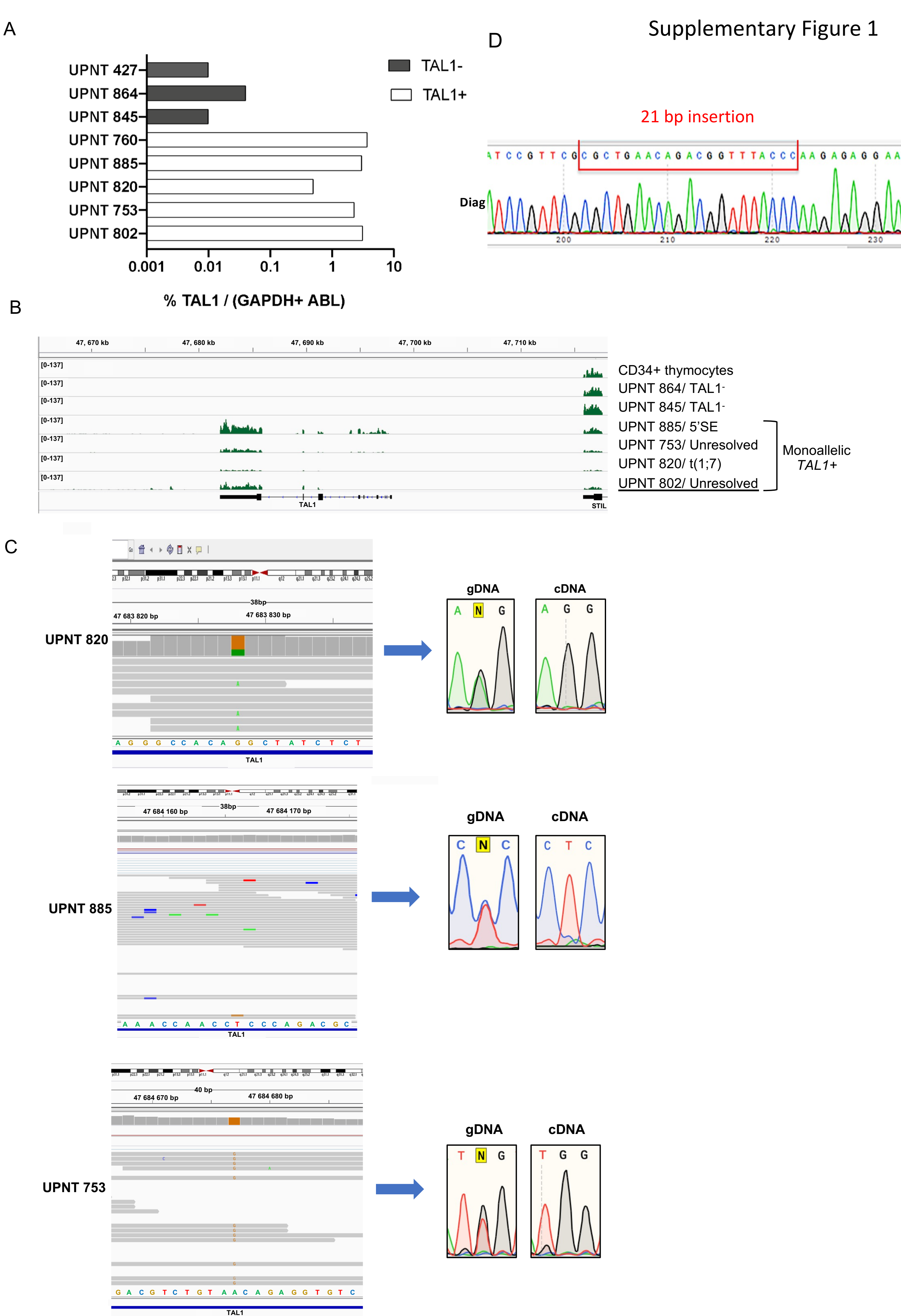
Name	Purpose	Sequence (5' to 3')	Note
Oligo 1 Tn5ME-A	Load adapter oligonucleotides	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	
Oligo 2 Tn5ME-B	Load adapter oligonucleotides	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	
Tn5MErev	Load adapter oligonucleotides	[phos]CTGTCTCTTATACACATCT	Complementary oligonucleotide
Tn5mCP1	Library PCR primers	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTC	
Tn5mCBar[1]	Library PCR primers	CAAGCAGAAGACGGCATACGA GAT(8-9N)GTCTCGTGGGCTCGG	Barcode[1] of 8 or 9 bases

[1]Barcode sequences: 1) GGATGTTCT, 2) CTTATCCAG, 3) GTAAGTCAC, 4) TTCAGTGAG, 5) CTCGTAATG, 6) CATGTCTCA, 7) AATCGTGGA, 8) GTATCAGTC, 9) TCGCCTTA, 10) CTAGTACG, 11) TTCTGCCT, 12) GCTCAGGA, 13) AGGAGTCC, 14) CATGCCTA, 15) GTAGAGAG, 16) CCTCTCTG, 17) AGCGTAGC, 18) AGCCTCG, 19) TGCCTCTT, 20) TCCTCTAC.

Copyright notice for barcode sequences 9 to 20: Oligonucleotide sequences © 2007-2012 Illumina, Inc. All rights reserved.

Supplementary Table 4: Predicted TAL1 complex binding motifs at 3'NE mutated sequences or within the vicinity using JASPAR prediction algorithm. The predicted binding motifs are underlined and are in bold. The mutation sequences are in red.

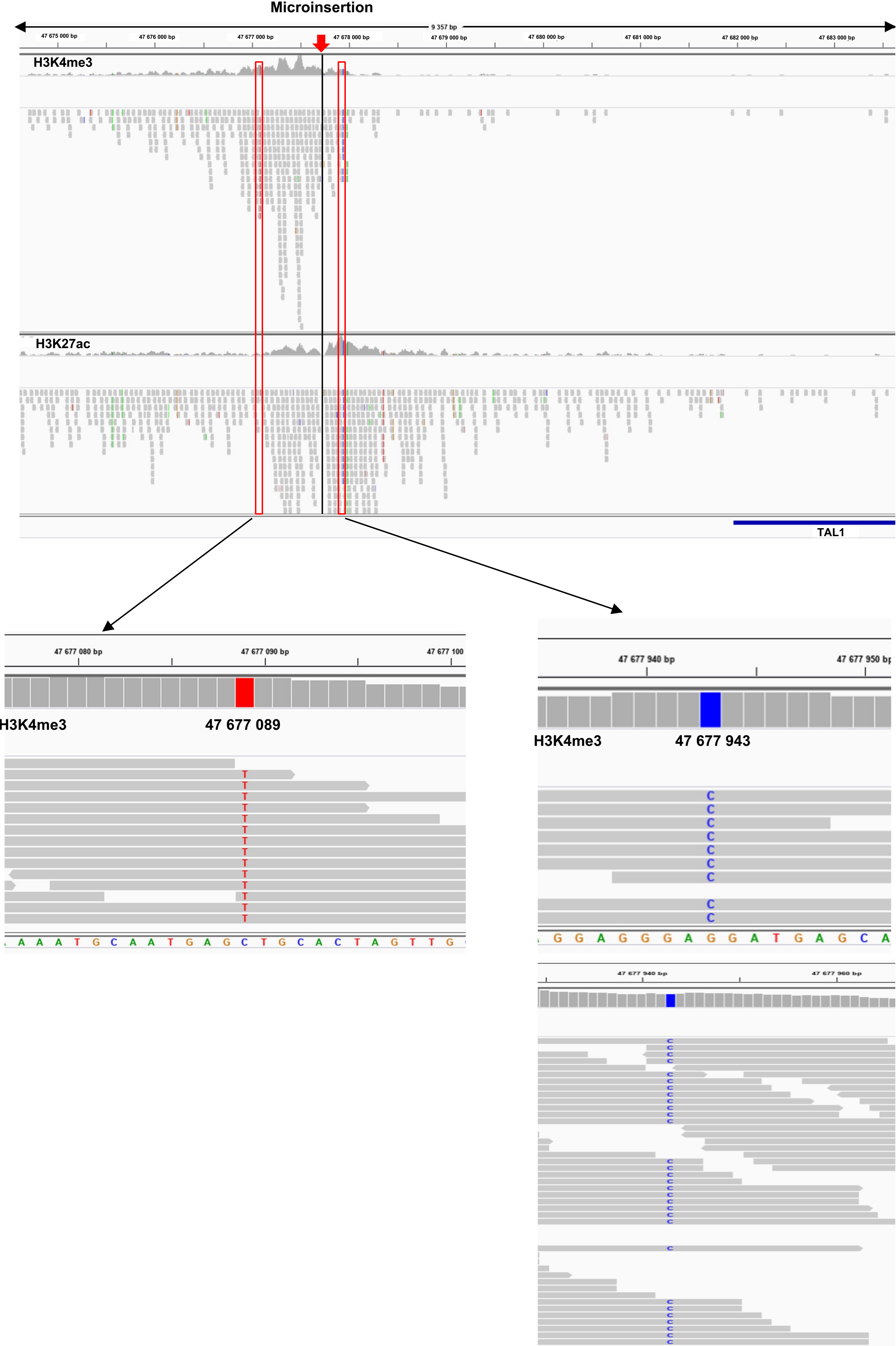
Patient	Protein	Predicted Sequence	Predicted Motif
UPNT802	MYB	GTGATCCGTTCGCGCTGAACA**GACGGTTT**ACCCAAGAGAGG	
UPNT802	GATA3	GTGATCCGTTCGCGCTGAACAGACGGTTTACCCA**AGAGAG**G	
UPNT802	TAL1	GTGATCCGTTCGCGC**TGAACAGACGGT**TTACCCAAGAGAGG	
UPNT802	RUNX1	ACACAATAACA**GGATGTGATCC**GTTCGCGCTGAACAGACGGTTTACCCAAGAGAGG	
UPNT613	MYB	GTGATCCGT**TAACGTTC**GAAGAGAGG	
UPNT613	GATA3	GTGATCCGTTAACGTTCGA**AGAGAG**G	
UPNT613	RUNX1	ACACAATAACA**GGATGTGATCC**GTTAACGTTCGAAGAGAGG	


Supplementary Figure 1
A
UPNT 427
UPNT 864
UPNT 845
UPNT 760
UPNT 885
UPNT 820
UPNT 753
UPNT 802
TAL1-
TAL1+
0.001
0.01
0.1
1
10
% TAL1 / (GAPDH+ ABL)
B
47, 670 kb
47, 680 kb
47, 690 kb
47, 700 kb
47, 710 kb
CD34+ thymocytes
UPNT 864/ TAL1-
UPNT 845/ TAL1-
UPNT 885/ 5'SE
UPNT 753/ Unresolved
UPNT 820/ t(1;7)
UPNT 802/ Unresolved
Monoallelic TAL1+
TAL1
STIL
C
UPNT 820
47 683 820 bp
47 683 830 bp
38bp
gDNA
cDNA
UPNT 885
47 684 160 bp
47 684 170 bp
UPNT 753
47 684 670 bp
47 684 680 bp
40 bp
D
21 bp insertion
Diag
200
210
220
230

Supplementary Figure 1: *TAL1* expression in primary samples

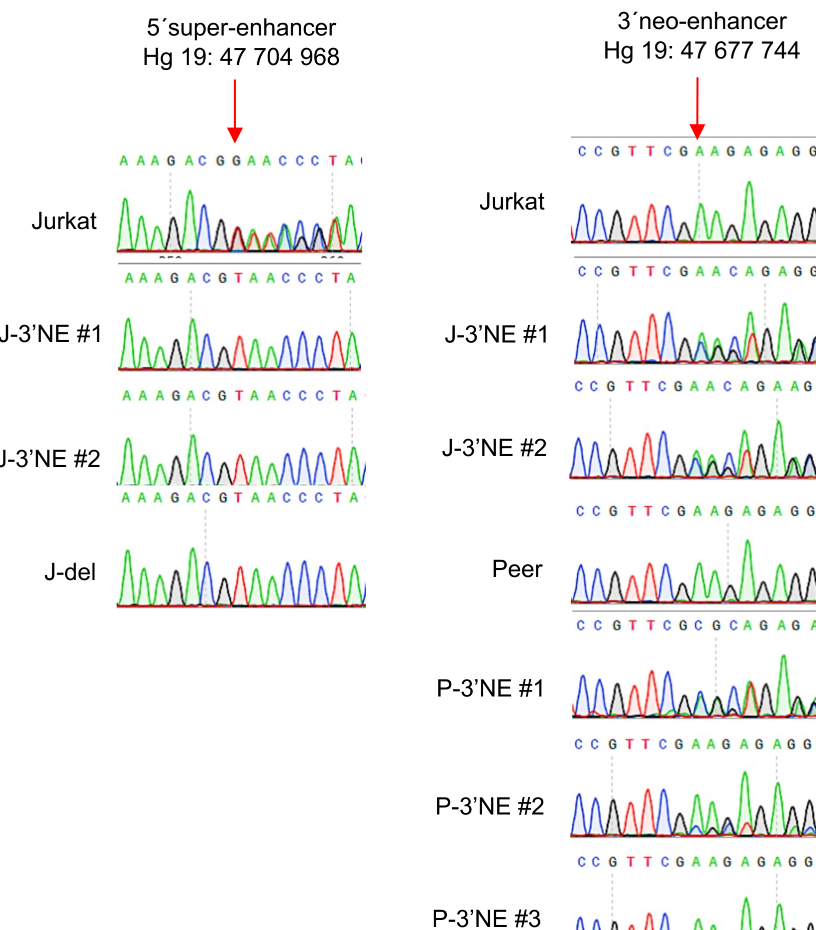
(A) *TAL1* expression measured by RQ-PCR and normalized to *GAPDH* and *ABL* housekeeping genes in T-ALL primary samples. (B) RNA-seq tracks centered over the *TAL1* locus in T-ALL primary samples where available. (C) *TAL1* allelic expression assessed by RNA-seq (left panels) or by Sanger sequencing of informative TAL1 3'-UTR SNP from gDNA or cDNA (right panels) in three TAL1 positive primary samples (UPNT-820, UPNT-885, UPNT-753). (D) Sanger sequencing chromatogram using microinsertion specific primers to better highlight the mutated sequence in the tumoral diagnostic sample.

Supplementary Figure 2



Supplementary Figure 2: (A) ChIP-sequencing track reads showing monoallelic enrichment of H3K4me3 and H3K27ac at two different SNPs in patient sample UPNT-802. (B) Zoom of ChIP-sequencing reads of patient sample UPNT-802.

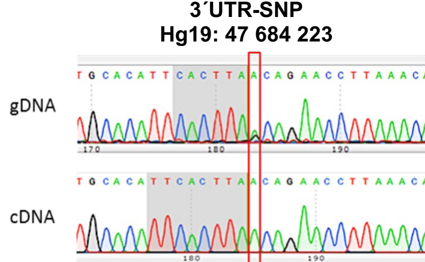
Supplementary Figure 3



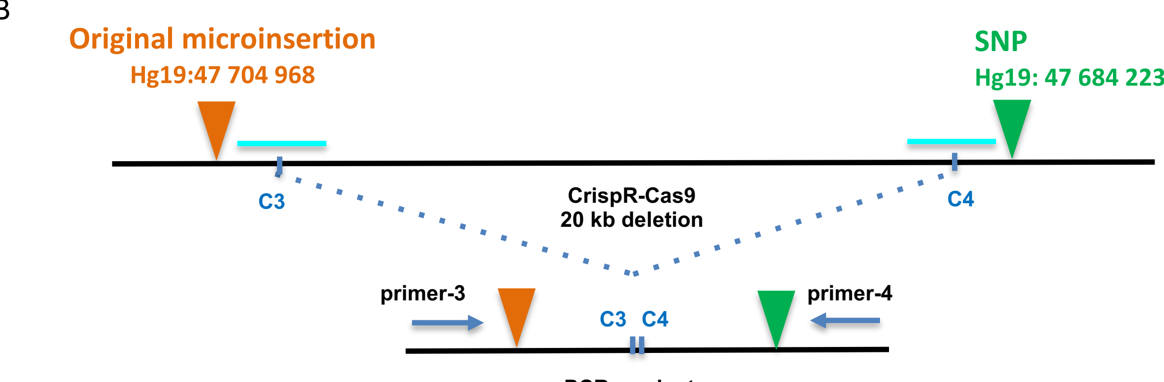
Supplementary Figure 3: Sanger genotyping of Crispr-Cas9 engineered cell lines

(A) Sanger chromatograms centered on the 5'microinsertion present in Jurkat and deleted in J-del, J-3'NE #1 and J-3'NE #2. (B) 3'microinsertion from UPNT-802 introduced in J-3'NE #1, J-3'NE #2, P-3'NE #1, P-3'NE #2 and P-3'NE #3 derivative cell lines.

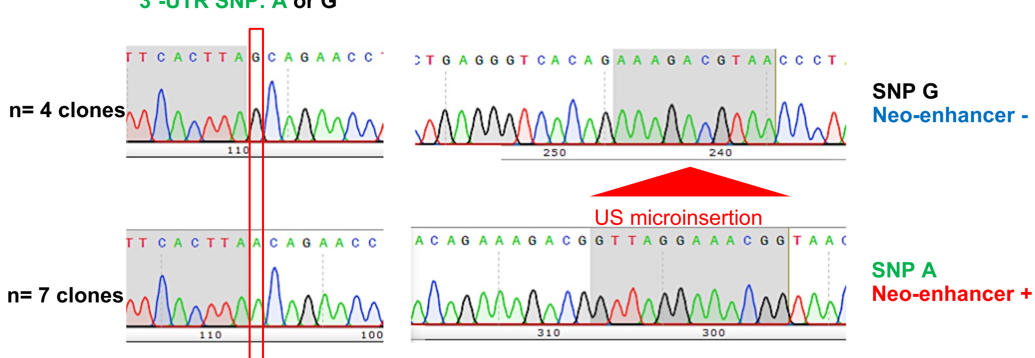
Supplementary Figure 4



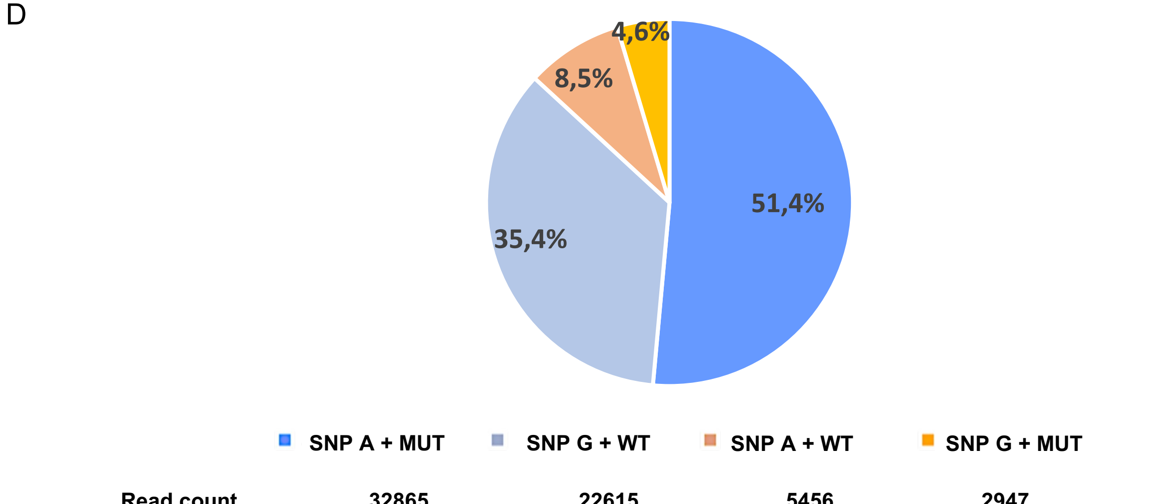
B



C



D

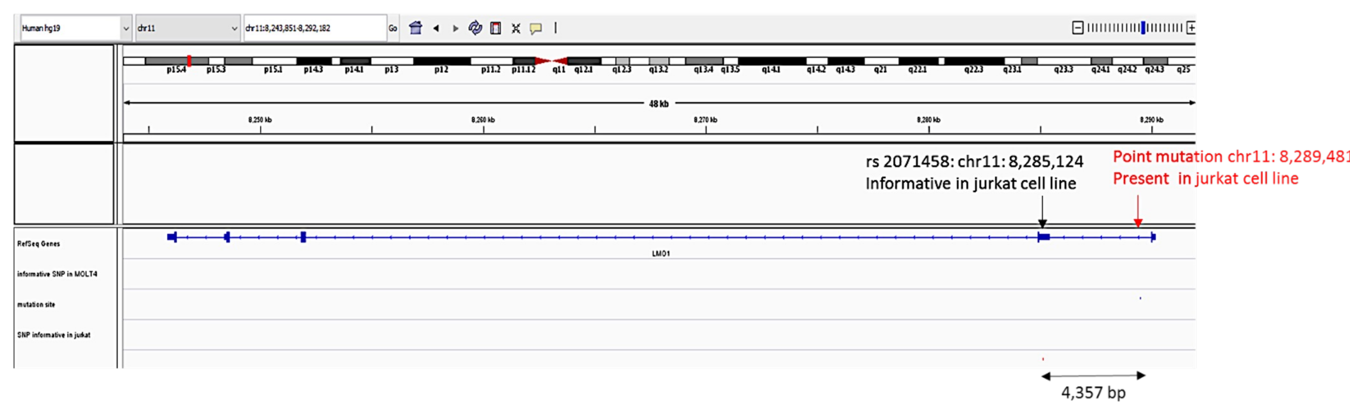


Supplementary Figure 4: 5'super-enhancer allele phasing

(A) *TAL1* allelic expression analysis in Jurkat cell line. Sanger sequencing chromatograms of the informative SNP in 3' UTR of *TAL1* from genomic DNA (gDNA; upper panel) or complementary DNA (cDNA; bottom panel) made from DNase treated RNA. Red box highlights the SNP. Allele A is expressed. (B) Schematic representation of the allele-phasing "Alpha-Cas" method. (C) After Cas9 deletion, the region containing the 3' UTR SNP and the 5'microinsertion was PCR-amplified and the PCR product was cloned in pBluescript SK+ plasmid. Representative chromatograms of 11 plasmid sequences are shown. 4 clones showed SNP G (left) and WT 3'sequence (right), 7 clones showed SNP A and mutated sequence. (D) The PCR product was also barcoded and sequenced with Mi-seq. Pie chart represents the respective fractions of the different phasing. A majority (86.8%) of reads show expected phasing (light blue and blue) with expressed SNP A phased with the microinsertion and the non-expressed SNP G phased with the WT sequence. A minority of reads (13.1%) show artefactual allele exchange (orange and beige).

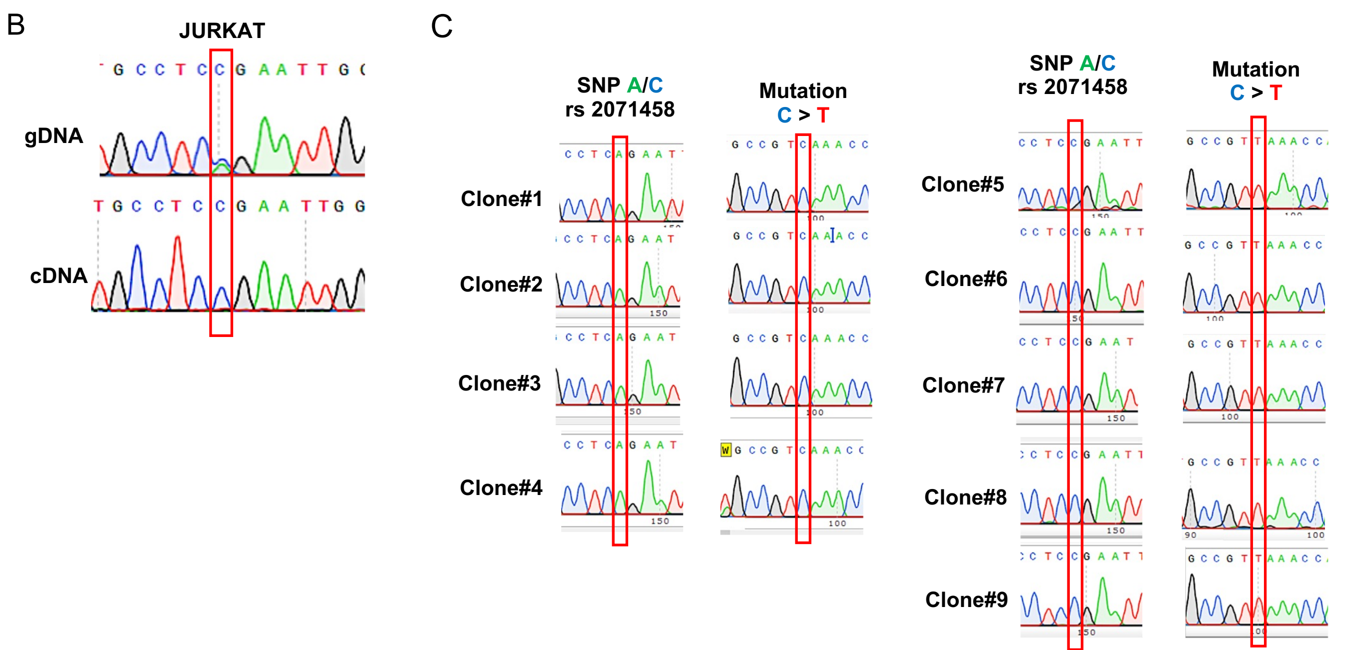
A

LMO1



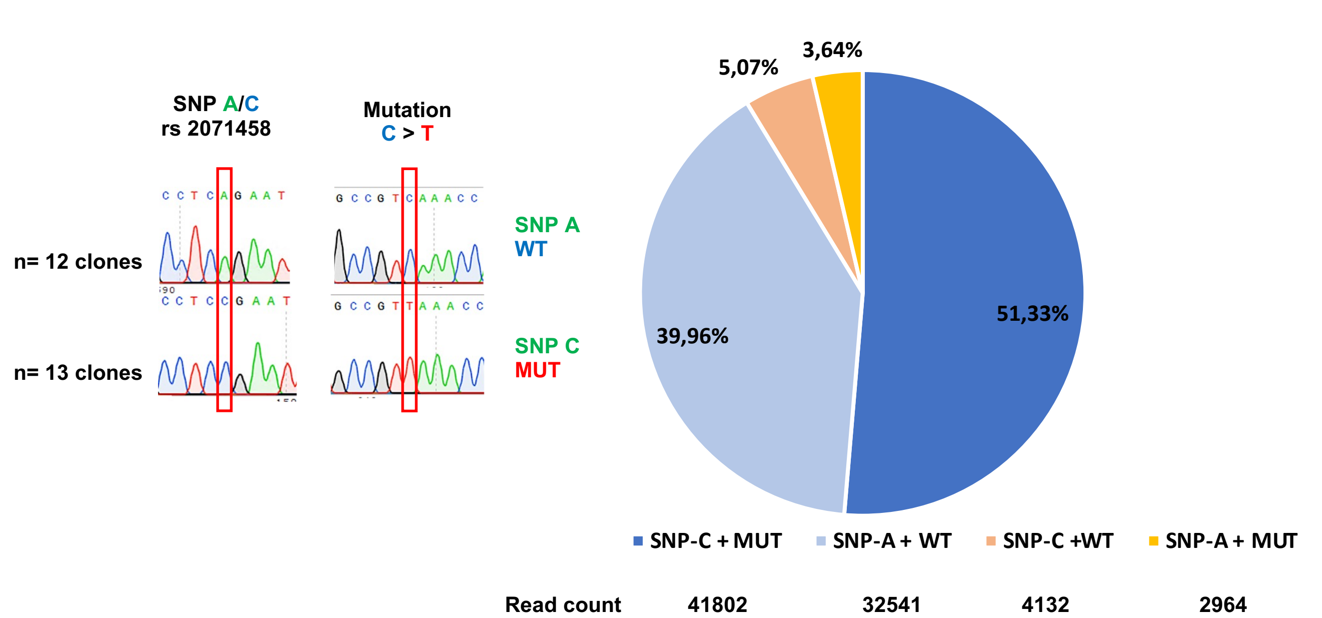
B

C



D

E



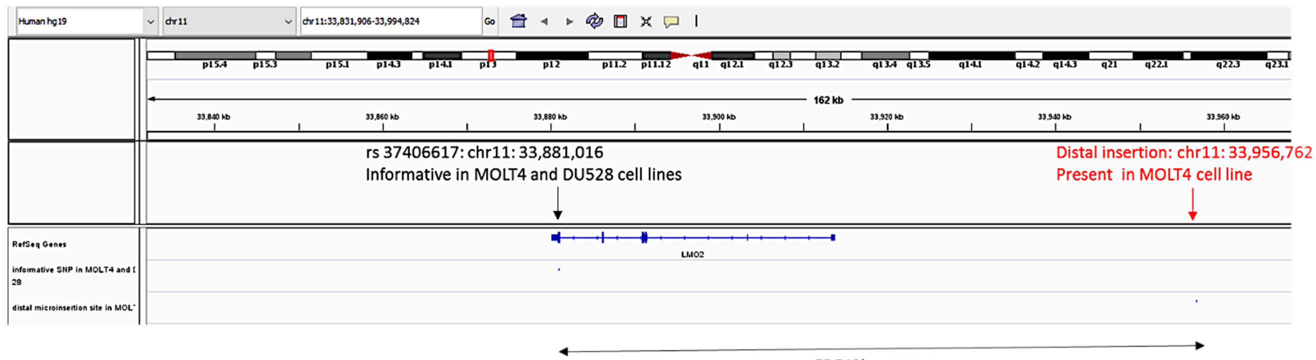
Supplementary Figure 5: *LMO1* mutation allele phasing

(A) LMO1 locus representation in Jurkat cell line. Black arrow highlights the heterozygous SNP (rs 2071458) and the red arrow highlights heterozygous point mutation in Jurkat cell line. (B) *LMO1* expression is monoallelic in Jurkat. Sanger sequencing chromatograms of the informative SNP from genomic DNA (gDNA; upper panel) or complementary DNA (cDNA; bottom panel) made from DNase treated RNA. Red box highlights the SNP. Allele C is expressed. (C) A 4357 bp genomic region including the informative SNP and the point mutation was PCR amplified and cloned in a pBluescript SK+ plasmid. Sanger sequencing chromatograms of 9 different clones are shown. 4 clones demonstrated the non-expressed SNP A together with the WT sequence (left panels); 5 clones demonstrated the expressed SNP C together with the mutated sequence (right panels). (D) After Cas9 deletion, the region containing the informative SNP and the mutation was PCR-amplified and the PCR product was cloned in pBluescript SK+ plasmid. Representative chromatograms of 25 plasmid sequences are shown. 12 clones showed SNP A (left) and WT sequence (right), 13 clones showed SNP C and the mutated sequence. (E) PCR product was also barcoded and sequenced by Mi-seq. Pie chart represents the respective fractions of the different phasing. A majority (91.29%) of reads show expected phasing (light blue and blue) with expressed SNP C phased with the mutation and the non-expressed SNP A phased with the WT sequence. A minority of reads (8.71%) show artefactual allele exchange (orange and beige).

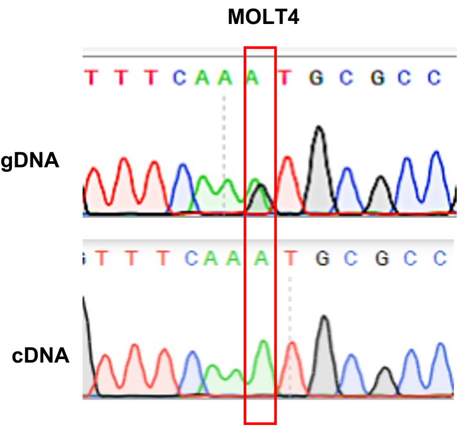
Supplementary Figure 6

A

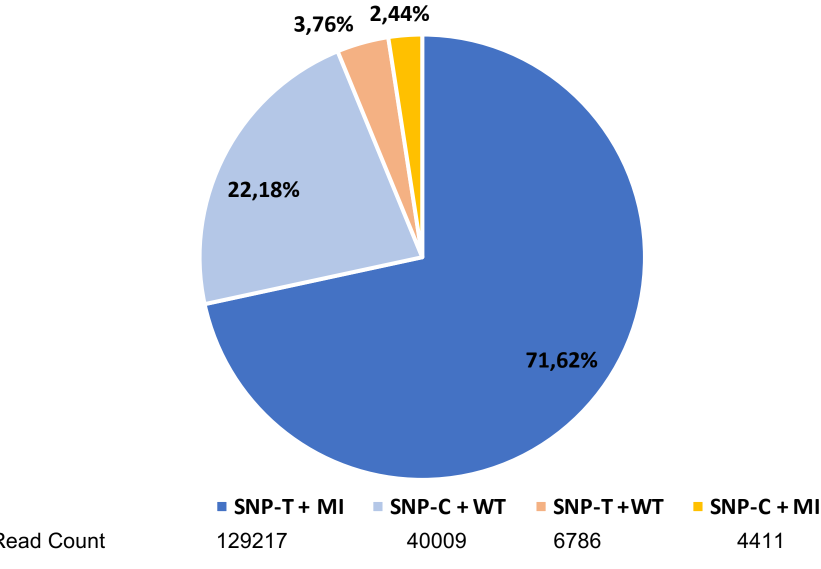
LMO2



B



C

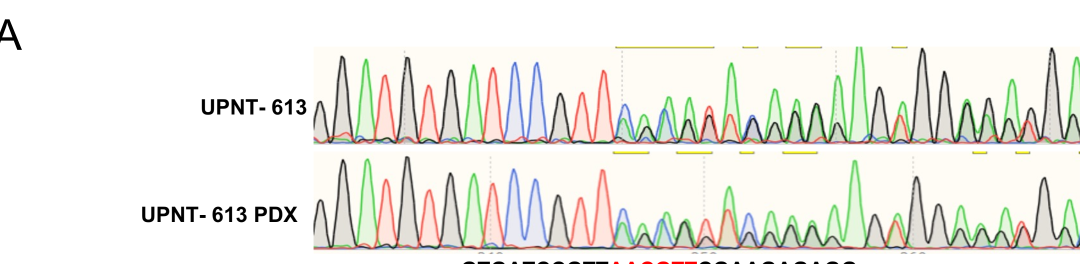


Supplementary Figure 6: *LMO2* mutation allele phasing

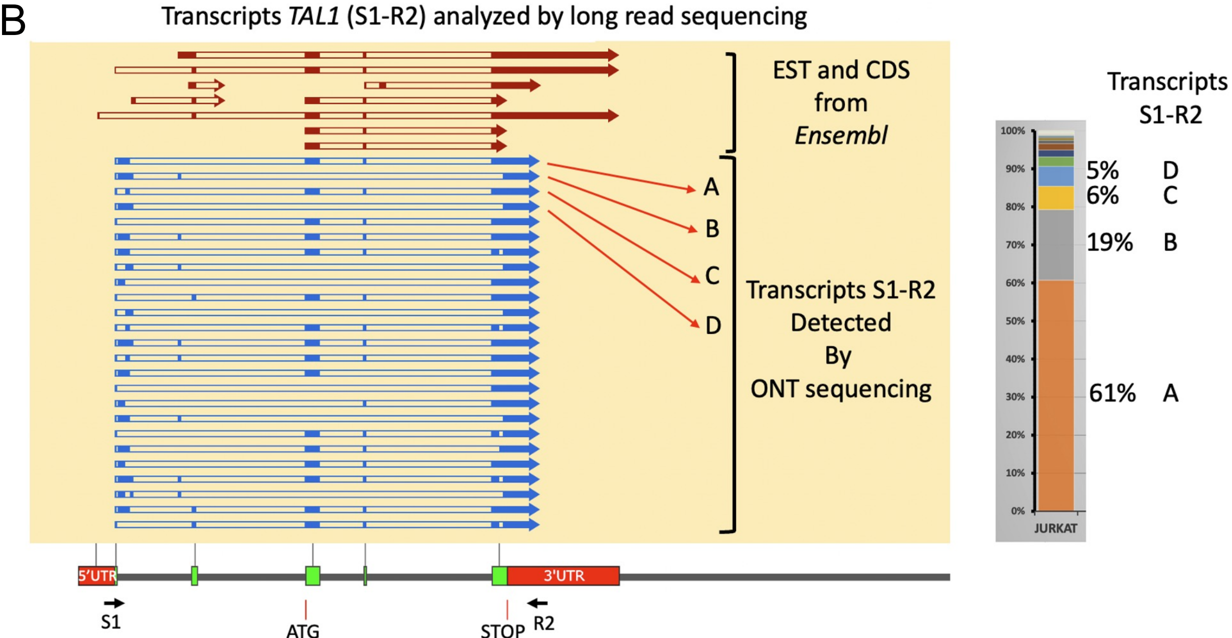
(A) *LMO1* locus representation in Molt4 cell line. Black arrow highlights the heterozygous SNP (rs 37406617) and the red arrow highlights the heterozygous microinsertion in Molt4 cell line. (B) *LMO2* expression is monoallelic in Molt4. Sanger sequencing chromatograms of the informative SNP from genomic DNA (gDNA; upper panel) or complementary DNA (cDNA; bottom panel) made from DNase treated RNA. Red box highlights the SNP. Allele A is expressed. (C) After Cas9 deletion, the region containing the informative SNP and the distal microinsertion was PCR-amplified and the PCR product was barcoded and sequenced by Mi-seq. Pie chart represents the respective fractions of the different phasing. A majority (93.8%) of reads show expected phasing (light blue and blue) with expressed SNP A phased with the microinsertion and the non-expressed SNP C phased with the WT sequence. A minority of reads (6.2%) show artefactual allele exchange (orange and beige).

Supplementary Figure 7

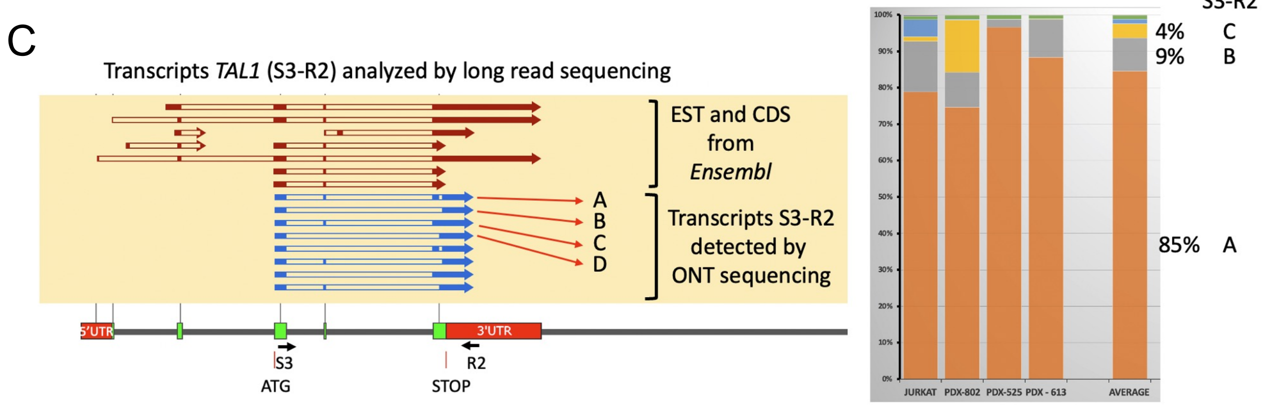
A



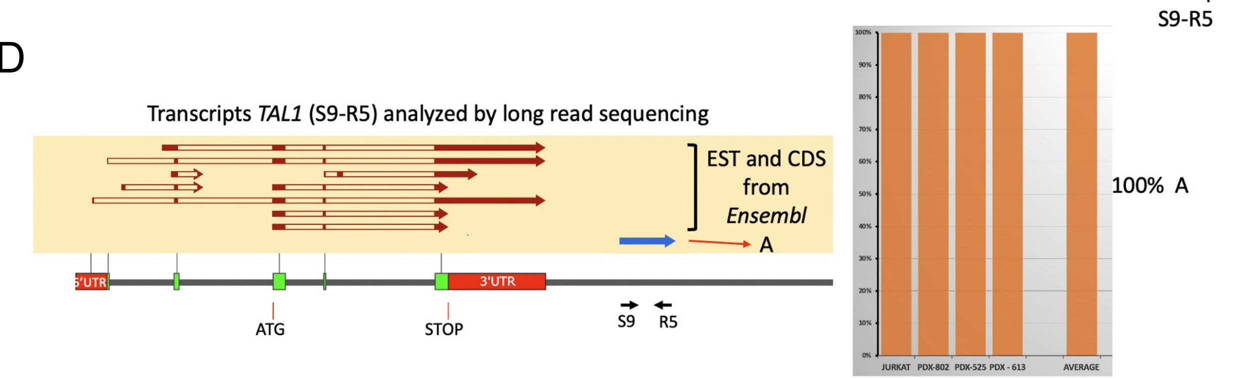
B



C



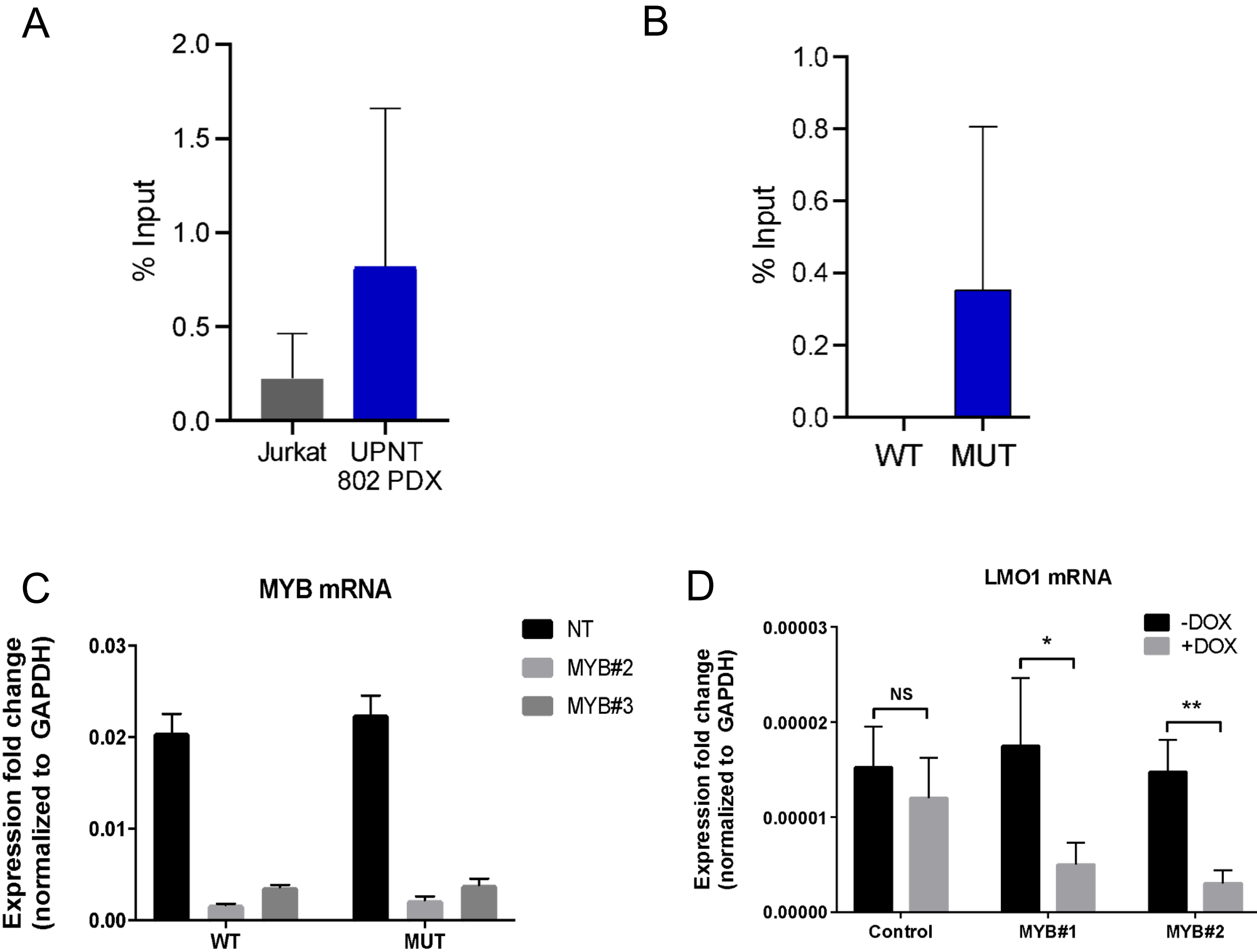
D



Supplementary Figure 7: Chromatograms of UPNT-613 diagnostic and PDX, and ONT analysis of TAL1 expression.

(A) A 6 bp microinsertion was found in UPNT-613 diagnostic and PDX samples downstream of *TAL1*. Sanger sequencing chromatograms are shown. (B-D) Schematic showing the alternative transcripts detected by long-read sequencing of several amplicons amplified by several primer pairs (S1-R2, S3-R2, S9b-R5). Expressed Sequence Tags (EST) and coding sequencing (CDS) as referenced in the Ensembl database are represented in dark red (Top left hand side panel), while the transcripts detected by ONT sequencing are represented in blue (Left hand side bottom panel). The distribution of the most abundant alternative transcripts is shown (Right hand side panel).

Supplementary Figure 8



Supplementary Figure 8: MYB mediated epigenetic activation at the at the 3'NE in 3'mutated PDX cells

(A) ChIP-qPCR showing MYB enrichment at the 3'NE in Jurkat and UPNT 802 PDX in duplicate. The mean and the SD are shown. Data was normalized to input controls. (B) ChIP-qPCR of MYB enrichment at the at the 3' WT sequence and 3'mutated sequence in UPNT 802 PDX in duplicate. The mean and the SD are shown. Data was normalized to input controls. (C) As verification of successful knockdown of MYB in jurkat cells, *MYB* expression was measured by RT-qPCR and normalized to *GAPDH* in the luciferase reporter assay. The mean and SD are shown of three replicates. (D) RTqPCR of *LMO1* expression after induced MYB knockdown as a positive control of successful knockdown in J-3'NE#1 cells.

Supplementary References

1. Ferrando A, Herblot S, Palomero T, et al. Biallelic transcriptional activation of oncogenic transcription factors in T-cell acute lymphoblastic leukemia. Oncogene 2004;103(5):1909–1911.
2. Navarro J-M, Touzart A, Pradel LC, et al. Site- and allele-specific polycomb dysregulation in T-cell leukaemia. Nat Commun 2015;66094.
3. Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat Biotechnol 2016;34(3):339–344.
4. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods 2014;11(8):783–784.
5. Sanson KR, Hanna RE, Hegde M, et al. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. Nat Commun 2018;9(1):5416.
6. Allemand E, Ango F. Analysis of Splicing Regulation by Third-Generation Sequencing BT - Alternative Splicing: Methods and Protocols. In: Scheiffele P, Mauger O, editors. New York, NY: Springer US; p81–95.